

Footstep localization based on in-home microphone-array signals

B. Van Den Broeck^{a,c,d} L. Vuegen^{b,c,d} H. Van hamme^b M. Moonen^{a,c}
P. Karsmakers^{a,c,d} B. Vanrumste^{a,c,d}

^a*ESAT-SISTA, KU Leuven, Kasteelpark Arenberg 10, 3001, Heverlee, Belgium*

^b*ESAT-PSI, KU Leuven, Kasteelpark Arenberg 10, 3001, Heverlee, Belgium*

^c*iMinds, Future Health Department, Kasteelpark Arenberg 10, 3001, Heverlee, Belgium*

^d*MOBILAB, TM Kempen, Kleinhoefstraat 4, 2440, Geel, Belgium*

Abstract. This paper describes a system able to detect footstep locations. Through acoustic information retrieved from a wireless sensor network with small and relatively cheap microphone arrays. A dataset was recorded in order to validate the accuracy of the detection. Results on this dataset show that a best median of errors of 31cm per time moment are achievable, but results heavily depend on the positions of the microphones relative to the footsteps.

Keywords. footstep location estimation, steered response power, global coherence field, gait analysis

Introduction

It is shown that a person's gait is related with there health, e.g. for early detection of dementia [1]. This work focuses on detecting footstep positions for later estimation of gait related parameters (i.e. such as walking speed, direction, stride length) by acoustic information, acquired from relatively cheap electret microphones. Acoustic sensing has the advantage that it is passive (i.e. no transmission of reference signals is needed) and contactless. For this work the setup of a wireless acoustic sensor network (WASN) has been chosen which differs from work using only one microphone array [2]. Such networks contain multiple so called nodes each holding at least one microphone. These WASNs have advantages over other kinds of setups. For instance the nodes can be small while maintaining large spatial sampling. The nodes can be placed uniformly in a room without inconvenient cables, which is ideal in an home environment. The computational load (which can be significant) can be distributed among nodes so that cheaper hardware can be selected [3].

This paper starts off with a general system description (section 1) and a selection of appropriate state-of-the-art algorithms for each part (section 2). Then an experimental setup and the data collection are described in section 3. In section 4, results on this dataset are presented. Finally conclusions are drawn in section 5.

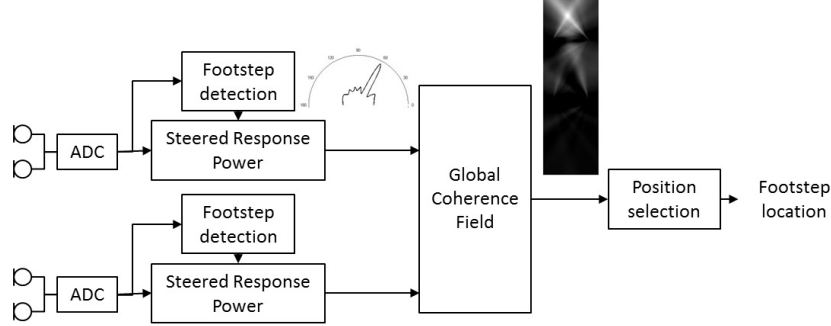


Figure 1.: Block schematic with example intermediate results.

1. System architecture

The proposed system consists out of 4 linear microphone arrays, further referred to as nodes. For esthetical reasons, these nodes should be rather small. In this work, nodes with 3 electret microphones with an inter-sensor distance of 6.8 cm are used, experimentally determined as a trade-off between beamwidth and node size. These nodes are then placed at ground level with a negligible microphone height of 4.5cm, making the localization a 2 dimensional problem. Furthermore, all node positions and orientations are considered to be known. A visualization of all tasks to be performed along with examples of intermediate results is shown in Figure 1. A first stage detects whether or not the sound contains a footstep. Through other alternatives have been studied for this purpose [4], this work uses an energy threshold for footstep detection. In a second stage each node is capable of estimating the sound energy coming from a certain direction. Multiple techniques exist but here the steered response power (SRP) is selected for reasons explained in section 2.1. The third stage combines the directional energies of all nodes to a map representing the spatial energy in 2D space. This is done by a global coherence field (GCF), further discussed in section 2.2. The last stage selects one XY coordinate out of the 2D GCF map, a step which is further discussed in section 2.3.

2. Methods

2.1. Direction of arrival estimation

Since sound travels at a finite speed, information about the direction of arrival (DOA) can be found in the time differences of arrival (TDOA) between different microphones in a node. The most simple way of measuring these TDOA is by a cross correlation, but this approach has a time resolution of only one sampling period. This problem can be resolved by using the so-called steered response power (SRP) algorithm [5]. SRP is based on a well-known delay and sum beamformer. To obtain time resolution the microphone signals are first chopped in frames (8000 samples which corresponds to 0.25 s at the used 32kHz sample rate, overlapping by 7000 samples), then the beamformer is steered in multiple directions at once (ranging 180° with 1° resolution). In each direction the retrieved energy of a frame is measured. In this work an enhancement on SRP is

used, namely SRP phase transform (SRP-PHAT). PHAT basically pre-transforms the microphone frames to have an unity spectral density. This operation decorrelates the different microphone signals over time, making the directional energy peaks narrower. The SRP-PHAT algorithm is further described in [5].

2.2. Global cohenrence field

Global coherence field (GCF) is a method for combining the directional energies of all nodes to a map containing energies at different XY positions. Unlike others, GCF takes all directional energy into account and not just the direction of the maximum energy. The main idea is to define an XY grid (points edging by 2 cm) and project the directional energies of all nodes onto this grid. This way intersections of high-energy directions will stand out. GCF is further described in [6]. Unfortunately, the collected data contains noise correlated over all microphones (presumably due to electromagnetic interference on the low level electret microphone output). But since these power contributions are quite stationary the noise effects can be estimated by a (local in time) average of the GCF map taken on noise only frames as follows:

$$GCF_{average}(k) = (1 - \lambda)GCF(k) + \lambda GCF_{average}(k - 1) \quad (1)$$

Subtracting this average map from the current map removes effect of the noise. The parameter λ (ranging from 0 to 1) symbolizes the time on which the average is taken. Prior tests show that $\lambda = 0.95$ (corresponding to a time constant of 20 ms) produced good results.

2.3. Source position selection

The GCF produces a map indicating the sound energy at different XY coordinates. The last step is to select one XY point out of this map as the footstep position. This is done by taking the weighted mean position of all points exceeding a threshold of 95% of the maximum, neighboring the maximum.

3. Experimental setup and data collection

The experiments were performed on data recorded in an office setting. Two test subjects were asked to walk 8 times on a predefined path containing 8 steps, yielding a total dataset of 128 (=2x8x8) steps. The nodes (plus front directions) and the footstep positions are shown in Figure 2a.

4. Results and discussion

Figure 2b shows boxplots of the errors for each footstep position. The median of errors range from 31 to 92cm. The figure indicates 2 important effects. Firstly that the errors are smaller for footstep positions enclosed by the nodes. This was expected since both signal-to-noise-ratio (SNR) and signal-to-reverberation-ratio (SRR) are larger and indicates the importance of a well chosen setup of the nodes. Secondly it is seen that the lower 50

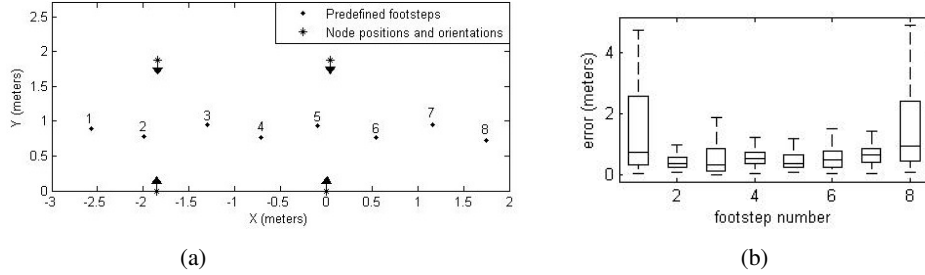


Figure 2.: a) Node positions/orientations and predefined footstep positions. b) Median of errors per footstep position.

percentile of errors is more dense than the upper 50 percentile indicating outliers. This is an important conclusion for future design of a post processor. Furthermore, following notes are important to contextualize these results:

- 1 This experiment is conducted on one floor type using only 2 types of footwear (2 test subjects), limiting the acoustics a footstep produces. The proposed system should be able to handle all sorts of footsteps acoustics, but the presented results can not be generalized.
- 2 The results presented here are based on a small set of footsteps. A larger data collection should be made to accurately validate the system.

5. Conclusions

This work focused on estimation footstep locations using acoustic information retrieved from a WASN. A system flowchart was presented and state-of-the-art algorithms were selected for each subtask. A dataset was recorded. The minimum and maximum observed error over the 16 instances per footstep position number are shown by the boxplots. Subsequent these experiments showed that the proposed system is able to detect footsteps with a best median of errors of 31cm. But also showed that these results are only achievable when nodes are located nearby the footstep location, making a well-chosen setup of nodes necessary. The experiment also showed that there are outliers in the position estimates for a frame, which will be of uppermost importance in designing a post-processing algorithm.

Acknowledgments

This work was preformed in context of following projects: ALADIN (IWT-SBO project contract 100049), IWT doctoral scholarships (contract 111433 and 121565) and FallRisk. The iMinds Fall-Risk project is cofunded by iMinds (Interdisciplinary Institute for Technology), a research institute founded by the Flemish Government. Companies and organizations involved in the project are COMmeto, Televic Healthcare, TP Vision, Verhaert and Wit-Gele Kruis Limburg, with projectsupport of IWT.

References

- [1] J. Verghese, C. Wang, R.B. Lipton, R. Holtzer, X. Xue, *Quantitative gait dysfunction and risk of cognitive decline and dementia*, Journal of Neurology, Neurosurgery & Psychiatry, vol. 78, pp. 929-935, 2007
- [2] M. Shoji, *Passive Acoustic Sensing of Walking*, Midori-cho, Musashino-shi, Tokyo, Japan, 2011.
- [3] A. Bertrand, *Applications and trends in wireless acoustic sensor networks: a signal processing perspective*, in Proc. of the IEEE Symposium on Communications and Vehicular Technology (SCVT), Ghent, Nov. 2011.
- [4] M. Tanaka, H. Inoue, *A study on walk-recognition by frequency analysis of footsteps*, Trans. IEE of Japan, vol. 119-C, No. 6, pp. 762-763, 1999.
- [5] I. Tashev, *Sound Capture and Processing*, John Wiley and Sons, 2009.
- [6] A. Brutti, M. Omologo, P.G. Svaizer, *Oriented global coherence field for the estimation of the head orientation in smart rooms equipped with distributed microphone arrays*, Eurospeech, Lisboa, 2005.